MONTHLY WEATHER REVIEW

Editor, ALFRED J. HENRY

Vol. 55, No. 6
W. B. No. 928

JUNE, 1927

CLOSED AUGUST 3, 1927
ISSUED AUGUST 30, 1927

A GROUP OR CORRELATION PERIODOGRAM, WITH APPLICATION TO THE RAINFALL OF THE BRITISH ISLES

By DINSMORE ALTER

[University of Kansas, Lawrence, Kans., April 22, 1927]

Recently, over a period of several years, the writer has been applying the Schuster periodogram to an analysis of the rainfall of typical sections of the world. These analyses of long records have been published from time to time, usually in the MONTHLY WEATHER REVIEW (1). At the very completion of the last of this series of papers he conceived the idea of a different periodogram, so simple in theory that it seems incredible that it can have escaped all investigators in this field. Earlier discovery of the method would have saved him hundreds of hours of computation.

An examination, both from theory and from an extended application to the rainfall of the British Isles, leads the writer to believe that the method contains the following advantages over the Schuster method:

(a) About one-third the calculation will permit a closer net to be spread than is usually done with the Schuster periodogram.

(b) Peaks often are located more definitely and interference cuts them away more rapidly than in the former.

(c) Examination for reality can often, indeed in general, be made in groups instead of individually.

(d) In many cases it is possible to discriminate regarding the physical reality of peaks of small amplitude. This has been impossible previously.

In the following discussion time will be regarded as the independent variable, although it is obvious that any other physical variable which has a functional relationship with a second might be used to advantage in certain physical problems. Suppose that we have equidistant observed functions of the time, $q_1, q_2, q_3 \ldots . q_n$. Let $q_j = q_{i+p}$; where p is the number of datum time intervals between the two q's. The data may be thought of as forming a curve with q's as ordinates and times as abscissæ. Let us match sections of the curve so that one section has been displaced by p time intervals. We shall now compute the Pearsonian correlation coefficient of these two sections and its probable error. These are:

$$r = \frac{\Sigma(\sigma_i \sigma_j)}{\sqrt{\Sigma(\sigma_i)^2 \Sigma(\sigma_j)^2}}; \quad \epsilon_r = 0.6745 \frac{1 - r^2}{\sqrt{\mu}}; \quad h = \frac{r}{\epsilon_r}$$

i refers to values from the first section of curve, j to those from the second, σ_i or σ_j is the departure of a q from the mean value of its section of curve and μ is the number of pairs of q's to be compared. In conservative standard usage μ should not be less than about 75. The greater the value of μ, of course, the better and the more reliable are the results obtained. If the curves are identical, or displaced in q by a constant amount, r is one. If they are exactly opposite, r is minus one. If they have no relationship, r is zero. The value of h determines the probability of a real physical relationship, that of r how close this relationship is. h will be as large as plus three in one pair of curves out of each 48 by accident, as large as plus four in one pair out of 290. The preceding is, of course, a mere restatement of the Pearsonian coefficient theory.

It is well known, of course, that a number of students have used the correlation coefficient for the purpose of disclosing suspected periodicities in data, but these uses have been limited, it is believed, to the investigation of particular periods. The novelty of the present proposal is to apply the principles of correlation to the evaluation of correlation coefficients for a systematic number of trial periods; in other words, to construct a periodogram similar to the well-known periodogram of Schuster.

Let us now select successive integral values of p and compute r/ϵ_r for each of them. We may plot a periodogram with these ratios as ordinates and the p's as abscissæ. Unlike the Schuster periodogram, the peaks do not depend for their height on a mere sine curve of period p, but on it and on the reinforcement of all its submultiples. Strength of any one of these will give us a peak in the periodogram, and a large number of such, even though their amplitude be rather small, will reinforce each other and make their presence apparent. As soon as we pass any point of maximum in the periodogram the shorter submultiples will if present rapidly get out of phase and their interference will cause a negative correlation to follow quickly. If they are absent, the primary alone being present, the peak will be broad as in the Schuster periodogram and will be lower than it would have been for a primary of the same amplitude with the harmonics present. Since an increase in the number of products increases the ratio for any given correlation factor r, there will be a general increase in the heights of peaks for small values of p, if only short periods or if comparatively few long periods are present.

A good practical length of periodogram to compute is from a maximum value of p to give about 75 products down to one one-half that length. If individual short periodicities are desired a Fourier series may be computed very quickly for the principal peaks of the periodogram. Since angles, sines, and cosines are the same for each term of the series, we are spared the great labor of determining many independent low peaks and yet are in no danger of overlooking any period of large amplitude. The periodogram enables us to throw out large groups of such peaks without individual calculation, thus making an enormous saving of time. It also makes practicable the use of an harmonic analyser. The computation of the periodogram is very short. With slide rule and adding machine, experience has shown that for 90 products of pairs, the ratio, giving one point of the periodogram, may be computed in about 30 to 40 minutes. Of course the sums of the squares of the residuals for the different p's are formed differentially in such speedy calculation. A periodogram of 50 points may, therefore, be calculated in about 30 hours after the data are secured. In the majority of problems the investigator will not be interested in the individual harmonics, but rather in their combined relationship, the calculation thus ending after these few hours of computation.

If one wishes to extrapolate future values from the facts secured through the data in hand it is unnecessary for him to make any calculation of phase angles unless he so desires for other purposes. Suppose that the periodogram has shown that there are certain periods x, y, z, etc., present, either as sine curves or with their harmonics. The investigator may make for each a table of the data, the first with x columns and as many rows as the data permit even using data which will not complete a row. He will take the average value in each column. These x means repeated over and over will give him the best representation of his data that he can make from the period x and its submultiples. Incidentally it will give the best prediction possible from these periods. He will then do the same for y, z, and any other peaks which he believes worthy of consideration. The results from all the periods considered are now averaged and reduced to the proper scale. In averaging, each period is given a weight equal to the number of cycles it has completed for the phase of that date. Amplitudes of the various terms are taken care of automatically by this method when we multiply the representation by the factor to adjust its mean deviation to that of the data and need not be computed for this purpose. Difficulties can arise only from submultiples which are common to two or more peaks. In general such will not be serious, but if concern is felt about them they may be subtracted easily from the final result by any of several quite obvious methods.

The accuracy of representation of the past data may easily be compared with that to be expected through accident. Let x be the number of peaks used in the representation and let y be the total number of datum values used by these peaks in representing any datum q; the quantity q has itself entered into the average, which is to represent it, x times, but, if there be no real relationship, the other $(y-x)$ points enter as accidental errors so that their sum increases as the square root of their number and equals statistically $\pm\sqrt{y-x}\epsilon$ where ϵ is the probable error of any one value of q. The sum of the y points is, therefore, $xq\pm\sqrt{y-x}\epsilon$ and the best adjustment to the q is made by dividing this by x, giving $q\pm\frac{1}{x}\sqrt{y-x}\epsilon$. The actual deviations of the representation are now found and their probable error p. The ratio H defined by the equation $H\equiv\frac{x}{\sqrt{y-x}}\frac{p}{\epsilon}$ is the factor by which residuals of the representation have been reduced from what would have been expected through mere accident and, therefore, gives an indication of the chances of successful prediction.

An investigation of the data for periods which are subject to systematic variations may be made by means of the same devices used for the Schuster Periodogram (*1a*). The limitations in this case are the same for each.

The grouping of the primary and its harmonics makes it possible, often, to investigate in a stretch of data periods which are longer than could be found if the sine curves were determined separately. Also, the rapid interference of the submultiples makes determination of their position more certain. If such shorter periods be not present we are able to carry periodicities to the lengths permissible in the Schuster method, though no farther. We have, however, even here, made a considerable saving of time.

It is obvious that if a peak in the periodogram exists mainly because of its even harmonics the periodogram will show a still higher peak at half that value, but that if it exists because of the odd harmonics there will be a large negative correlation found at that point. All harmonics contributing would tend toward a zero correlation at the half point.

As an example of the application of the method, data which already had been used in an application of the Schuster Periodogram and given as British Isles B in the tables of a previous paper (*1f*), have been analysed. The Schuster analysis had brought out a number of short periods, which because of their correspondence in periodograms computed from chronologically different stretches of data appeared to be real. All of these had been shown to be very closely harmonics of a period between 43 and 43½ years but closer to the former. This periodogram is computed for data lags of 24 to 52 years. The period of 43 years found by the Schuster method appears prominently but is not the most prominent one. The reason for this reversal is that the former calculation was made from data beginning with 1850, British Isles A of the earlier paper (*1f*). These have been extended back to 1834 and an additional station added. The first half of 1848 was unusually wet and matched with a dry half year later to give almost the maximum possible negative product, one so large that even with more than 90 products of pairs it was serious. Had it not been for this one-half year of very early data the 43-year peak would have overshadowed all the others.

We find in the periodogram three peaks which are more than three times the probable error. They are at lags of 24½, 41, and 51 years. It was noticed that these bear very simple ratios to each other, i. e., ⅗ and ⁵⁄₄, respectively, to the middle one. The ratios are far more exact then the distance between computed points. This simple relationship suggested the finding of the least common multiple, in order that it might be determined whether the principal ones of the lower peaks were also harmonics. This multiple is 613 years. The fourth and fifth highest peaks are the 43 plus and the 34 year Brückner cycle. The former is represented at 43.8 by the fourteenth harmonic, a little to one side of the crest but exactly at the center of the base of this peak, and the latter is perfectly represented by the eighteenth. There are a number of peaks of low enough height that none by itself can carry any weight as being at all more than mere accident. The great success of representation of the highest peaks suggested the idea of computing all the harmonics in the range covered by the periodogram. The accuracy of the additional representation is surprising.

Of all the peaks in the periodogram only one fails to be demanded by a harmonic and only one harmonic, the twenty-fourth, does not fit a peak of some, although many times negligible, height. The two highest of these minor peaks are represented perfectly by the thirteenth and seventeenth harmonics. In only 3 out of the 14 consecutive harmonics is the error as large as 10 per cent of the interval between it and the next nearest peak. It seems that these peaks can not be due to some one, or to a very few short periods which are harmonics of all. Undoubtedly that will account for part of the representations, but it can not for all of them.

The conclusions are that the fit is not accidental and that it can not be due to some variable cycle. The simplest explanation, though not, of course, the only possible one, is that these terms are exactly what they seem—harmonics of an irregular period of about six centuries. Whether this be the true explanation is impossible to know at the present time.

The four principal peaks are all within accidental amounts of the same height. Low peaks need, of course, not be considered in representation. Where to stop is a matter of judgment only. These four have been combined in the manner described earlier in the paper, and the accuracy of the representation is shown by Figure 2. In 88 per cent of the data the deviations are on the same

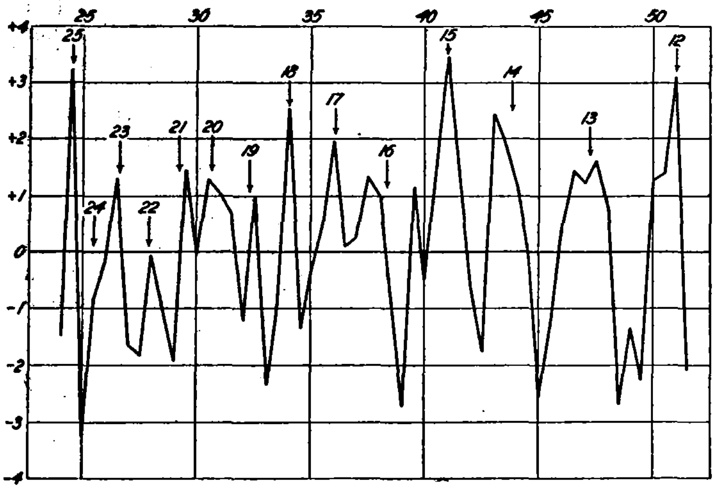

FIG. 1.—Correlation periodogram of rainfall British Isles 1834 to 1924.

side of the normal. The correlation factor between the representation and the data is +0.91.

The computation of the probability of such close representation emphasizes it still more accurately. Using the notation of a preceding paragraph, we find that for the 182 values represented, y equals 9 in 38 cases, equals 10 in 108, equals 11 in 20, and equals 12 in 16 cases. We may therefore consider y as equal to 10 throughout the whole of the curve. x equals 4, the number of peaks used.

We compute $p=4.6$; $\epsilon=10.5$; $H=0.71$. The reduction in size of the residuals from that demanded by error theory to 71 per cent of that value is great enough that we may feel rather confident that the correlation between the extrapolated part of our curve and the data yet to be observed will be positive. Whether it will be sufficiently large to have other than academic interest is a matter on which the writer will hazard no opinion. *The future values are given for test purposes only.* Since they result from the combination of four cycles, they should carry more weight than the test values given in the previous paper, which depended entirely on but one of these.

The writer wishes to acknowledge the grant from the research committee of the University of Kansas under which a large part of the computing was done and the excellent service of Mr. James O. Edson, who was engaged to perform it.

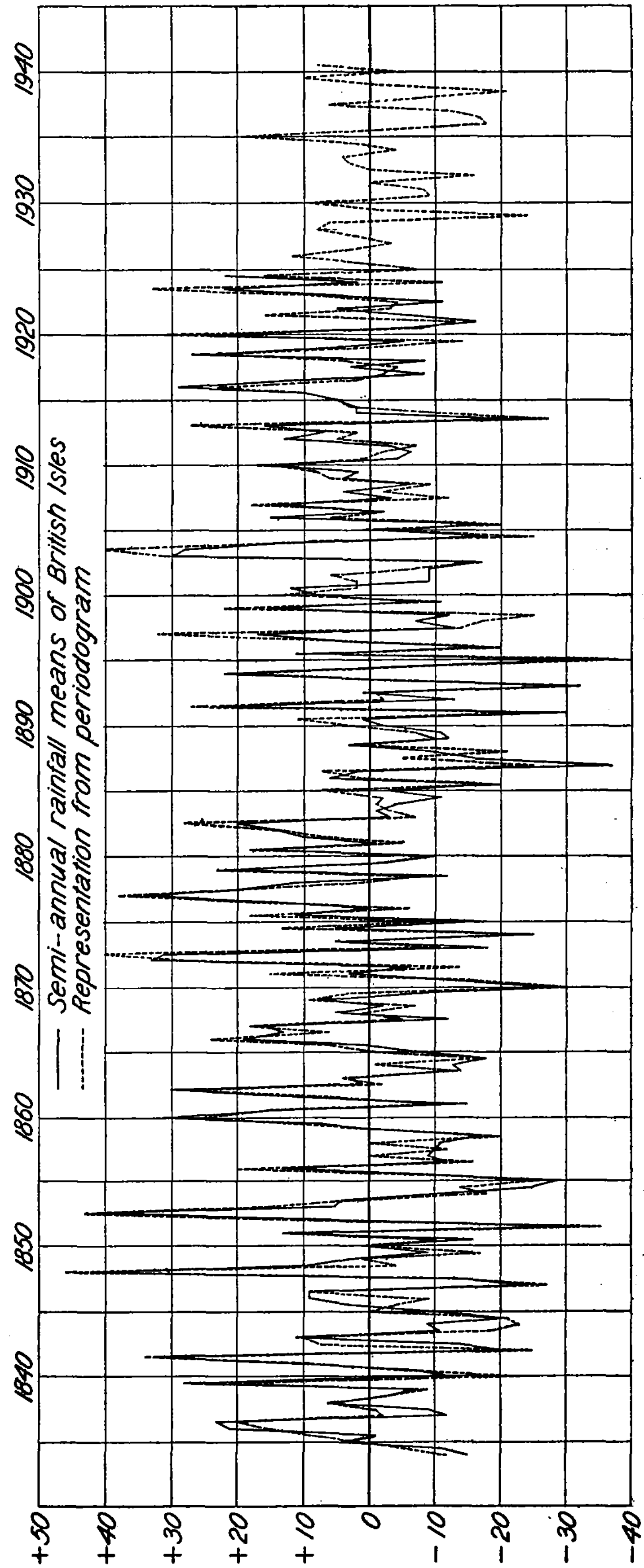

FIG. 2.—Representation from periodogram of semiannual rainfall of British Isles

LITERATURE CITED

(1) ALTER, DINSMORE.
(a) AN EXAMINATION BY MEANS OF SCHUSTER'S PERIODOGRAM OF RAINFALL DATA FROM LONG RECORDS IN TYPICAL SECTIONS OF THE WORLD. Mo. WEA. REV., 54 : 44–56. February, 1926.
(b) THE CRITERIA OF REALITY IN THE PERIODOGRAM. Ibid. 57–58.
(c) APPLICATION OF SCHUSTER'S PERIODOGRAM TO LONG RAINFALL RECORDS BEGINNING 1784. Mo. WEA. REV., October, 1924, 52 : 479–487.
(d) THE SIXTEEN-YEAR PERIOD IN CLIMATE. Bulletin American Meteorological Society, August, 1925.
(e) EQUATIONS EXTENDING SCHUSTER'S PERIODOGRAM. Astronomical Journal, No. 850.
(f) 1927. INVESTIGATIONS OF RAINFALL PERIODICITIES BETWEEN 1½ AND 2½ YEARS BY USE OF SCHUSTER'S PERIODOGRAM. Mo. WEA. REV., February, 1927, 55 : 60–65.
(g) A STUDY OF THE POSSIBILITY OF ECONOMIC VALUE IN STATISTICAL INVESTIGATIONS OF RAINFALL PERIODICITIES. Mo. WEA. REV., March, 1927, 55 : 110–112.

SOLAR RADIATION OBSERVATIONS AT APIA, SAMOA

By ANDREW THOMSON, Director

[Apia Observatory, Samoa, May 19, 1927]

The Apia Observatory purchased a Gorczyński pyrheliometer in 1924 in order to carry on a series of measurements on the amount of solar radiation transmitted by the humid atmosphere of this Tropic island (long. 171° 46′ W. lat. 13° 48′ S.).

The pyrheliometer had been previously calibrated by Dr. H. H. Kimball with the United States Weather Bureau standard and found in good agreement with it. (1) There has been no opportunity to recalibrate since that time, but two facts indicate that the constants of the pyrheliometer, if carefully handled, do not change over considerable periods. (1) This instrument had been calibrated in Europe several years prior to the Washington calibration and after a strenuous expedition to Tougourt in the Sahara Desert had retained its accuracy, as shown by the Washington calibration, to an accuracy of 1 per cent. (2) The recorded values in 1927 are practically the same as those taken at the same season of the year 1925.

The base to which the thermopile was attached was rigidly fastened to a post in the center of a small platform erected on the seashore. The thermopile was about 2 m. above mean sea level. It was connected by 33.5 m. of the 16-gauge copper wire to the recording millivoltmeter in the main office. The exposure was good, but since the equatorial mounting had been made for the Northern Hemisphere the base had to be placed vertical instead of horizontal, causing a greater strain on the driving gear than that for which it had been designed. Probably on this account the driving clock did not keep the thermopile facing directly into the sunlight without frequent readjustment. Between the hours of 8 a. m. and 4 p. m. the instrument was carefully watched. The values of the solar radiation for air masses greater than 2 may be too small due to the thermopile not being properly oriented.

A table was prepared giving for each day of the year the instant when the air mass was 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, etc. Using this table, the records of the millivoltmeter were measured and the values at the times given in the table were tabulated. In the case of obvious irregularities in the record caused either by clouds or the thermopile having got out of line, a smooth curve was drawn through the dots when registration was satisfactory and the tabulated values derived therefrom. The present investigation includes only those days in which there were at least four hours' good record. Since rain squalls and overcast skies are of very frequent occurrence during the months November to March, the data for these months are rather scanty.

In Samoa the year may be divided into three seasons—*wet season* (November–February), 51 per cent rainfall, trade winds irregular; *equinoctial* (March–April, September–October), 33 per cent rainfall; beginning and end of trade-wind season; *dry season* (May–August), 16 per cent rainfall, trade winds strongly developed. During the dry season the solar radiation is approximately constant (1.1 gm. cal.) from 9 a. m. to 3 p. m. At the other seasons of the year this period of constant solar radiation is longer because the sun is farther south and remains at high altitudes for a longer period of the day.

In Table 1 under Q is given the observed solar radiation. This shows that in the wet season the radiation is more intense than in the dry season, and that when morning and afternoon values of radiation passing through

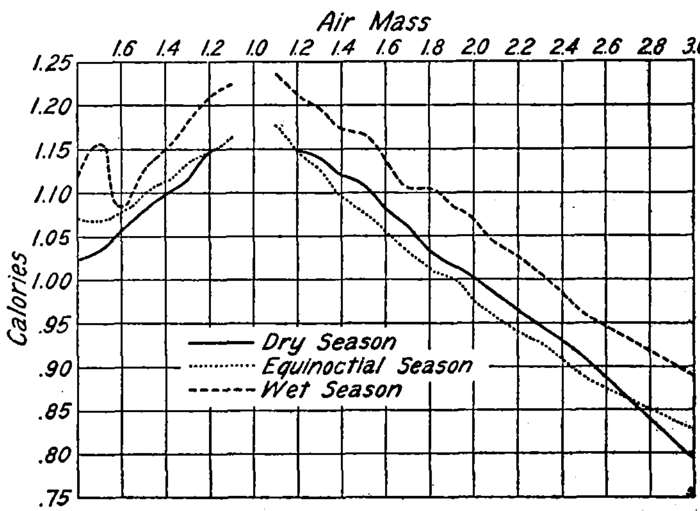

FIG. 1.—Solar radiation intensity, sat normal incidence, reduced to mean distance of the earth from the sun. (Gram-calories per minute per square centimeter)

the same air mass are compared during the wet and dry seasons the afternoon values are the greater, but in the equinoctial months the morning values are the greater.

During the dry season the earth is at a greater distance from the sun than in the wet season, so that before comparing the transparency of the atmosphere it is necessary to adjust the values of Q to unit distance. The dry season values were multiplied by 1.028, the wet season by 0.973, and the equinoctial values were unchanged. As shown by the values of Q' in Table 1 and Figure 1, the radiation values are uniformly higher in the wet season.